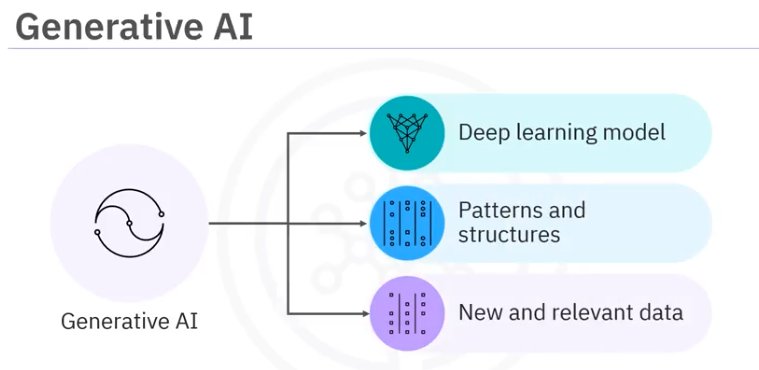
[Significance of Generative AI](https://www.coursera.org/learn/generative-ai-llm-architecture-data-preparation/lecture/SG0qi/significance-of-generative-ai#)

**Gen AI:** DL models used to generate high quality texts, images and other contents based on the data they were trained on. Trained to understand underlying trends and patterns within the data. Generative AI is like an artist who examines various paintings, understands the patterns, and produces original art inspired by what they learned. 

Text Generation: Generative narrative, GPT (Generative Pre-trained Trasformers)

Image Generation: Text to Input (DALL -E), From seed image or random input (GAN, Diffusion model)

Audio Generation: Natural sounding speech, text to speech synthesis Ex; Wavenet

**Application:**

* Content creation
* Condensing documents
* Lang translation
* Chatbots and virtual assistants
* Data Analysis

***According to Bloomberg Intelligence, generative AI is expected to become a one point $1.3 trillion market by 2032.***

This will expand its applications in various domains. Some examples are enhancing personalized recommendations, contributing to medical breakthroughs through drug discovery, and integrating generative AI into smart homes and autonomous vehicles.

[Generative AI Architectures and Models](https://www.coursera.org/learn/generative-ai-llm-architecture-data-preparation/lecture/ba1Df/generative-ai-architectures-and-models#)

RNN: Seq/time series data, loops in strcuture

Transformers: Feedback mechanism, self-attention mechanism, most important parts ex: GPT

GANs: Generator, discriminator, adversarial process

VAEs: Encoder, Decoder, ex; Art and creative design

Diffusion models: Statistical properties (distorted to clear images), remove noise and reconstruct

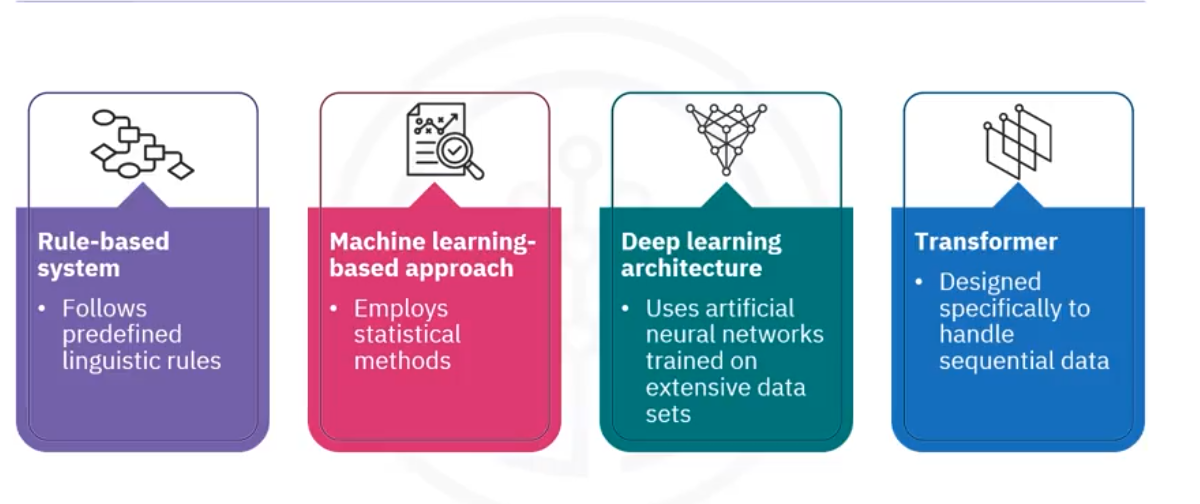


Related to reinforcement learning: learn from environment to get rewards.

[Generative AI for NLP](https://www.coursera.org/learn/generative-ai-llm-architecture-data-preparation/lecture/YbJqG/generative-ai-for-nlp#)

Applications: Chtabot (Air canada, medical, bank) , Text summarization (QuillBot (Summarizer Tool), TLDR This, LinkedIn (post/article previews), Slack (thread summary bots), Microsoft Word (Smart Summarize via Copilot)) , Machine Translation(Google translator), Sentiment Analysis(E-commerce platforms: Amazon)

**Scenario: AI Engineer in a bank**

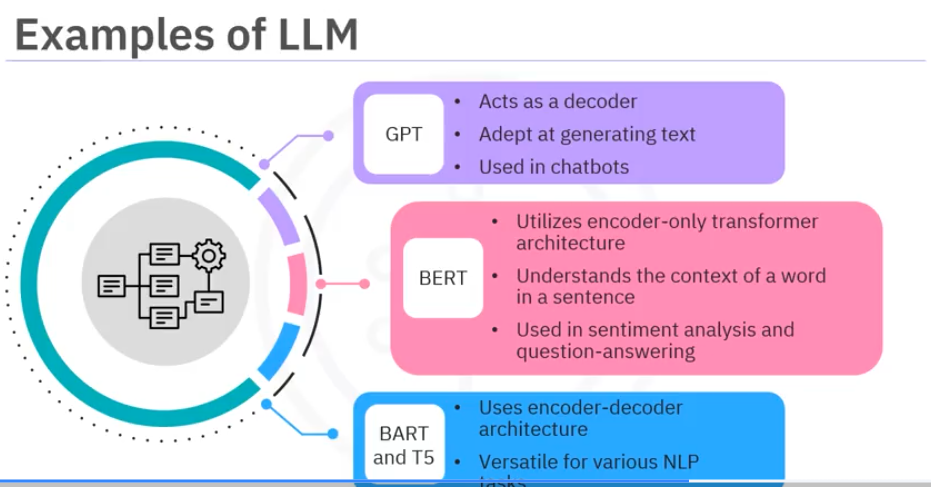
****

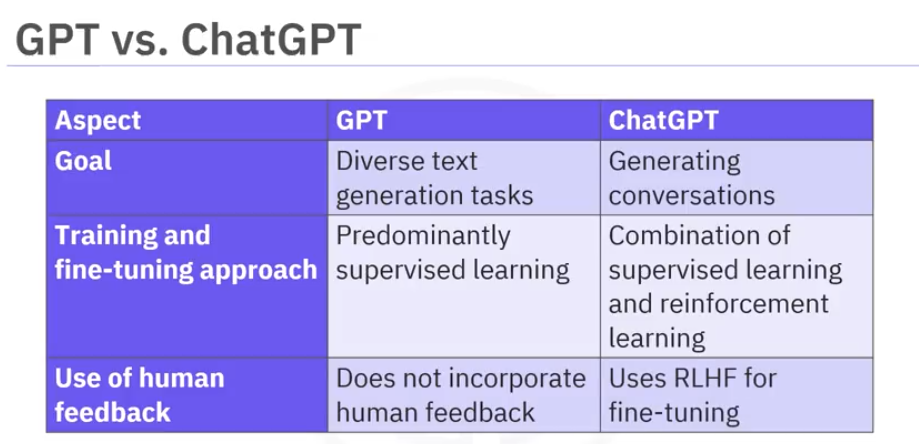
**LLM:**

1. Uses AI and DL with vast datasets
2. Involve training datasets running into petabytes
3. Billions of parameters which are fine-tuned

Models:

* pretrained transformer series or **GPT series.**
* Bi-directional encoder representation from transformers or **BERT,**
* Bi-directional and autoregressive transformers or **BART** and
* text to text transfer transformer, also called **T5**

****

****

**RLHF: Reinforcement Learning from Human Feedback**

### 

### **AI hallucinations**

In AI hallucinations, the model generates output that it presents as accurate but is seen as unrealistic, inaccurate, irrelevant, or nonsensical by humans. It is similar to the way humans experience hallucinations.

For example, there was an incident where ChatGPT falsely claimed that a mayor in Australia was found guilty and imprisoned in a bribery case. In reality, the mayor notified the authorities about a bribery issue. (Reference: [Australian mayor readies world's first defamation lawsuit over ChatGPT content | Reuters](https://www.reuters.com/technology/australian-mayor-readies-worlds-first-defamation-lawsuit-over-chatgpt-content-2023-04-05/))

**Reasons:**

Factors such as biases in the training data, limited training, complexity of the model, and lack of human oversight can cause AI hallucinations.

### **Methods for mitigating hallucinations**

* Eliminating any bias in the training data and performing extensive training of the models on high-quality data
* Avoiding manipulation of the inputs that are fed into the models
* Ongoing evaluation and improvement of the models
* Fine-tuning a pre-trained LLM on domain-specific data

### **Preventing the problems caused by AI hallucinations**

It is inevitable for hallucinations to occur within LLMs. What can be frustrating is that the generated text often contains subtle mistakes that are challenging to identify. There are a couple of best practices that you can follow. These include:

* Being vigilant and understanding that these models do not understand the actual meaning of the words but are focused on predicting the next word in a sequence based on patterns. These models are trained on vast amounts of data and learn statistical patterns, but they lack semantic understanding or comprehension **like human beings**.
* Ensuring human oversight regularly for fact-checking and continuous testing
* Providing additional context in the prompt or input. This will enable LLMs to understand the desired output better and generate more accurate and contextually relevant responses.

# Reading: Overview of Libraries and Tools

* Pytorch (Developed by meta research lab, used by OpenAI, Tesla)
* Tensorflow (Developed by google)
* Hugging Face
* LangChain
* Pydantic

**Text Generation before Transformers:**

1. N-gram language models
2. RNN
3. LSTM and GRU
4. Seq-to-seq

## 

## Transformers

Proposed in a paper titled "Attention Is All You Need" by Vaswani et al. in 2017, the transformer architecture replaced sequential processing with parallel processing. The key component behind its success? The attention mechanism, more precisely, self-attention.

Key steps include:

* Tokenization: The first step is breaking down a sentence into tokens (words or subwords).
* Embedding: Each token is represented as a vector, capturing its meaning.
* Self-attention: The model computes scores determining the importance of every other word for a particular word in the sequence. These scores are used to weight the input tokens and produce a new representation of the sequence. For instance, in the sentence "He gave her a gift because she'd helped him", understanding who "her" refers to requires the model to pay attention to other words in the sentence. The transformer does this for every word, considering the entire context, which is particularly powerful for understanding meaning.
* Feed-forward neural networks: After attention, each position is passed through a feed-forward network separately.
* Output sequence: The model produces an output sequence, which can be used for various tasks, like classification, translation, or text generation.
* Layering: Importantly, transformers are deep models with multiple layers of attention and feed-forward networks, allowing them to learn complex patterns.

The architecture's flexibility has allowed transformers to be used beyond NLP, finding applications in image and video processing too. In NLP, transformer-based models like BERT, GPT, and their variants have set state-of-the-art results in various tasks, from text classification to translation.

Lab: # \_\_Exploring Generative AI Libraries\_\_ ("facebook/blenderbot-400M-distill")

